

The Advantages of Using Machine Learning in Digitizing Manuscripts

General Engineering

Moumen Alkhateb Alhasani

201910156

Qatar University

ENGL 203 – 007

April 23, 2020

Muhammad Farrukh Ashraf Syed

The Advantages of Using Machine Learning in Digitizing Manuscripts

Digitizing manuscripts revives the glories of nations. No way to know the circumstances of the pioneers of scientific discoveries without returning to what was narrated and written about them in manuscripts. Challenges had always been behind exploring the contents of manuscripts until the computer science revolution and artificial intelligence made digitizing effortless through machine learning. Digitization is creating a digital soft copy of a hand-written document by scrutinizing the dark and light parts in order to identify each alphabetical letter or numerical digit and transmuting each character into an ASCII code. Some advantages of digitizing manuscripts are publishing a checked version, preserving the original, and making manuscripts more accessible.

The first benefit of digitizing manuscripts is eliminating transcription errors. The traditional steps to transcribe a manuscript are reading it word by word, rewriting it clearly, comparing the newly written version with the original, and submitting the checked copy to a printing house to be typed and published. While being copied, an undecipherable word or a complete sentence in some cases commonly appear in such a way that they are not legible due to the font complexity and letters overlapping. This problem is innovatively solved by digitization. The purpose of manuscript replication is fundamentally to make it readable (Madden & Seifi, 2011). Since digitization technology makes manuscripts clearer, it facilitates their readability, and this is due to the fact that computers are able to identify all shapes of letters and their irregular rules especially when they are combined, which is something that humans may not always be capable of achieving. This allows the amanuensis to focus on the comprehensible digitized form rather than the original manuscript, making the task much easier since digitization through machine learning eliminates these transcription errors. Another type of error is the typographical error. Before digitizing, the amanuensis' final copy was sent to a printing house to be entered into a computer and then published. However, those working with

the manuscript may not be sufficiently familiar with the subject matter. Consequently, a significant number of misinterpretations followed by serious mistakes may invade manuscripts during the publishing process. However, Optical Character Recognition (OCR) technology performs this operation smoothly and precisely, and its accuracy could reach an astonishing high percentage. It can classify each letter and determine manuscripts' language even if they contain more than one, not to mention its ability to distinguish between different fonts in a single language. OCR becomes more accurate by dealing with more data, and this is literally what machine learning is.

Furthermore, conserving manuscripts is an indispensable advantage of digitization. One type of threat to manuscripts comes from mishandling and human negligence. Examples of mishandling could be violent flipping, holding manuscripts with contaminated fingers, folding pages of manuscripts to mark where the reader stopped, and eating while reading. These malpractices seriously damage manuscripts, which may eventually lead to their total destruction. Nevertheless, a digitized manuscript will not be affected by such mishandlings as it is in a digital form. Hence, safeguarding manuscripts by diminishing frequent physical handling is one of the purposes of digitization (Qatar National Library, n.d.). In fact, in connection with mishandling manuscripts, digitization becomes more crucial for rare manuscripts. Allowing the use of such manuscripts might endanger their survival, while preventing access to them might be withholding important forms of knowledge. For that reason, digitization becomes the only rescue for manuscripts, which allows a balance to be struck between these two priorities. The ability of artificial intelligence in digitization is almost unbelievable. Charred, torn manuscripts with obscure letters, and shrunk fragile scrolls that are on the point of falling apart can finally be preserved using machine learning algorithms. Other types of mishandling are counterfeiting, theft, and destruction, like throwing countless precious manuscripts into the Tigris River by Mongols in 1258 that made the waters run black with ink.

If these books were available presently, machine learning technology would identify and digitize their content after analyzing high-energy rays passed through them, no matter how substandard they are. Moreover, digitization conserves manuscripts from environmental factors. The method of storing manuscripts determines their longevity (Meij, 2017). Manuscripts deterioration comes from several causes, such as dampness, fungi, dirt, contamination, temperature fluctuation, air pollution, bacteriological and radiological factors. If the standards of care and preservation are considered, not every manuscripts library is perfectly air-conditioned. Consequently, digitizing manuscripts is the unrivaled custodian that preserves them for future use.

Finally, digitizing manuscripts makes them easy to access. Before, people had to travel to places or even to other countries to get access to manuscripts and benefit from them. Even after people have traveled distances, they might be denied physical access to the manuscripts, and if they did get access, the permission would be limited by a specific time duration and uncomfortable restrictions. However, and as most institutions including universities, libraries, and museums share their digitized documents with the public gratis, manuscripts can be viewed easily at any time at the reader's convenience. The noteworthy benefit of digitization is to facilitate distant access to what institutions hold for people who may not be able to reach them physically (Prescott & Hughes, 2018). In other words, the accessibility of digitized manuscripts is for all people globally, while it is only for those who can reach their physical location before the advent of digitization. It is also worth mentioning that people used to depend on photographs of manuscripts for later references. As it is known, normal photography creates huge files depending on their qualities, but not every person can easily download them because of the different capabilities of computers and networks being used for this purpose. Besides, normal photography would not enable searching and selecting texts for copying or modifying as computers would consider it as an image, not a document. However, machine learning

technology converts manuscripts to a small, editable, and more accessible file, which means that more people could utilize them as need be. Moreover, digitization provides simultaneous access to multiple users to the same manuscript inclusively and swiftly without any impediment. This is a paramount advantage to schools and universities, especially for students of literary and historical studies, who are interested in ancient heritage. Traditionally, most, if not all, institutions prevent synchronous access to paper manuscripts and limit it to a tiny group of scholars. However, digitization technology augments the allowable limit to be infinity instead of having a very small number of users.

In conclusion, the benefits of machine learning in digitizing manuscripts are invaluable. Transcription or typographical errors cannot find a way to infiltrate the digitized form of manuscripts as digitization technology promotes document readability and eliminates human errors. Digitization also preserves manuscripts' content from improper handling by people and it rescues them from inferior ecological circumstances. Besides, digitizing manuscripts internationalizes their accessibility by providing unlimited concurrent access without having to be in the same physical location. With the development of artificial intelligence, it is believed that the capability of digitization technology will grow and diversify spectacularly so that ancient civilizations will continue to stay relevant for generations to come.

References

Digitization services. (n.d.). Retrieved from link <https://www.qnl.qa/en/library-services/digitization-services>

Madden, K., & Seifi, L. (2011). Digital surrogate preservations of manuscripts and Iranian heritage: Enhancing research. *New Library World*, 112(9), 452-465. doi: 10.1108/03074801111182049

Meij, D. (2017). *Indonesian manuscripts from the islands of Java, Madura, Bali and Lombok*. Available from link https://books.google.com.qa/books?hl=ar&lr=&id=YtMuDwAAQBAJ&oi=fnd&pg=PR7&dq=indonesian+manuscripts+from+the+islands+of+java+madura+bali+and+lombok&ots=MGNFm3_1Xx&sig=50i5BfgroqJsPe1ErgHHHEPU0X4&redir_esc=y#v=onepage&q=indonesian%20manuscripts%20from%20the%20islands%20of%20java%20madura%20bali%20and%20lombok&f=false

Prescott, A. & Hughes, L. (2018). *Why Do We Digitize? The Case for Slow Digitization*. Retrieved from link <https://www.archivejournal.net/essays/why-do-we-digitize-the-case-for-slow-digitization>



The First
Evidence.pdf



The Second
Evidence.pdf



The Third
Evidence.pdf



The Fourth
Evidence.pdf



SafeAssign
Report.pdf



Rubric.pdf